

Más allá del GWAS: alternativas para localizar QTLs

Filippo Biscarini, Stefano Biffani, Alessandra Stella *PTP (Parco Tecnologico Padano), Lodi, Italy*

April 2015

Se presentan dos métodos que se pueden utilizar como alternativas al GWAS para localizar QTLs en ganado. El primero se centra en el análisis de secuencias de homocigosis (runs of homozygosity, ROH) en grupos de animales (p.ej. casos y controles); el segundo es un método de remuestreo basado en la frecuencia de inclusión de los SNP en modelos predictivos. Las ROH se aplicaron a la identificación de regiones del genoma asociadas a trastornos reproductivos en vacas de raza Frisona, y el método de remuestreo a la detección de portadores del haplotipo BH2 en vacas de raza Pardo Alpina. Estos métodos alternativos pueden complementar el GWAS estándar en la localización de QTLs para caracteres de interés en el genoma de animales domésticos.

[English abstract]

Beyond GWAS: alternatives to localize QTLs in farm animals. Two methods that could be used for QTL mapping as alternatives to standard GWAS are presented. The first relies on the differential frequency of runs of homozygosity (ROH) in groups of animals (e.g. cases and controls), while the second stems from resampling techniques used for the prediction of carriers of a mutation, and is based on the frequency

of inclusion of polymorphisms (SNP) in the predictive model. ROH were applied to the detection of reproductive diseases in Holstein-Friesian cattle, while resampling was applied to the detection of carriers of the BH2 haplotype in Brown Swiss cattle. These alternative approaches may complement GWAS analyses in localizing more accurately QTLs for traits of interest in livestock.

Introducción

El descubrimiento de regiones del genoma asociadas a caracteres de interés zootécnico (p.ej. resistencia a enfermedades) es una importante etapa preliminar para posibles aplicaciones de la genómica a la cría de animales domésticos como la identificación de portadores de mutaciones favorables/desfavorables o la selección asistida por marcadores. Esto es ahora posible gracias a la gran cantidad de datos procedentes de las nuevas técnicas de secuenciación masiva (NGS: next generation sequencing) en forma p.ej. de genotipados de alta densidad de SNPs y de secuencias de genoma completos. Los estudios de asociación de genoma completo (GWAS) se han convertido por lo tanto en una técnica estándar para escanear el genoma en busca de polimorfismos asociados al fenotipo analizado. Sin embargo, los métodos de GWAS presentan algunas limitaciones: analizan gen-

eralmente un SNP a la vez, son susceptibles de producir resultados espurios y, salvo en caso de señales de asociación muy patentes, pueden resultar de difícil interpretación (McCarthy et al. (2008)) . En este artículo se presentan dos alternativas al GWAS para localizar QTLs en ganado: el análisis de secuencias de homocigosis (“runs of homozygosity”, ROH) y un método de remuestreo basado en la frecuencia de inclusión de los SNP en modelos predictivos.

Material y Métodos

Para ilustrar las diferentes propiedades de los métodos basados en GWAS, ROH y remuestreo para localizar QTLs, se utilizaron los resultados de tres estudios previos con bovinos lecheros de raza Frisona y Pardo Alpina. En particular: el GWAS se basó en 17 vacas afectadas por el desplazamiento abomasal izquierdo y 47 controles de raza Frisona (Biffani et al. (2014)); el análisis de ROH en 163 vacas afectadas por trastornos reproductivos y 295 controles de raza Frisona (Biscarini et al. (2014b)); y el método de remuestreo en 513 portadores del haplotipo BH2 en BTA19 y 3132 no portadores de raza Pardo Alpina (Biffani et al. (2015)). Todos los animales se genotiparon con el chip bovino 50k de 54 001 SNPs.

GWAS : El estudio de asociación pangenómico (GWAS) para el desplazamiento abomasal izquierdo se llevó a cabo siguiendo el procedimiento GRAMMAR-GC descrito por Aulchenko et al. (2007): primero, se ajustó un modelo de regresión logística para datos binarios (caso-control) incluyendo los efectos sistemáticos de granja, orden de parto y el efecto poligénico aleatorio, y posteriormente se utilizaron los residuos de este modelo en una regresión lineal para estimar el efecto de cada SNP.

Análisis de secuencias de homocigosis (ROH)

: Bajo la hipótesis de que las enfermedades complejas presentan una componente genética constituida por muchas variantes recesivas distribuidas a lo largo del genoma, se analizaron las secuencias de homocigosis (ROH) para identificar regiones asociadas a trastornos reproductivos en vacas lecheras. Las ROH se definen como secuencias de genotipos homocigotos adyacentes, que reflejan la transmisión de haplotipos idénticos por parte de progenitores comunes. En lugar de centrarse en un locus, las ROH consideran también al entorno, teniendo así en cuenta el cambio de frecuencia de SNP cercanos que estén en desequilibrio de ligamiento con el locus analizado. Se identi-

ficaron entonces las ROH presentes en la población de vacas bajo estudio a través de la estimación de la homocigosis de los SNP en “ventanas deslizantes” de 1000 kb. Se permitieron como máximo 5 genotipos faltantes y 1 genotipo heterocigoto para considerar la secuencia como ROH. La frecuencia relativa de las ROH en casos y controles indica regiones del genoma asociadas con el fenotipo.

Remuestreo de SNPs en modelos predictivos

: Para la identificación de portadores del haplotipo BH2 en vacas de raza Pardo Alpina se aplicó un procedimiento basado en la reducción progresiva del número de SNP incluidos en el modelo y en el análisis discriminante lineal (LDA). Para cada umbral de SNP utilizados (2.5%, 10%, 15%, 30%, 50% y 100% sobre el total de SNP) se seleccionaron aquellos más predictivos a través del método de Best Subset Selection (BSS) y se utilizaron los SNP seleccionados para clasificar los individuos en portadores o no del haplotipo. El procedimiento se repitió 1000 veces (validación cruzada de 10 pliegues, 100 réplicas) por cada umbral, y se obtuvo la gráfica de la frecuencia de inclusión de los SNP en el modelo predictivo en función de su posición en BTA19. De esta manera se pudo localizar la mutación en el genoma.

Resultados y Discusión

En el estudio de asociación se detectaron SNPs asociados con el desplazamiento abomasal izquierdo en BTA12. La asociación más significativa se halla a una distancia de tan solo 20 kb del gen *SLITRK5*. Este gen pertenece a la familia génica *SLITRK*, implicada en procesos neurológicos que ya se demostraron asociados a la patogénesis del desplazamiento abomasal en vacuno. La Figura 1 presenta el Manhattan plot del análisis GWAS. Se puede apreciar la falta de señales claras de asociación y la aparición de numerosas asociaciones, potencialmente espurias, que resultan más o menos “significativas” y distribuidas a lo largo del genoma.

La interpretación de los resultados puede ser arbitraria, lo que ilustra algunas de las limitaciones del GWAS. El análisis de las secuencias de homocigosis (ROH) permite ampliar la mirada al entorno de cada SNP, llevando a interpretaciones más robustas de los resultados: la aparición de un haplotipo homocigoto en casos pero no en controles puede ser más informativa que un SNP asociado individualmente. La Figura 2 ilustra el caso de trastornos reproductivos en vacas lecheras en BTA15. El análisis de ROH es un método falto de un modelo estadístico explícito,

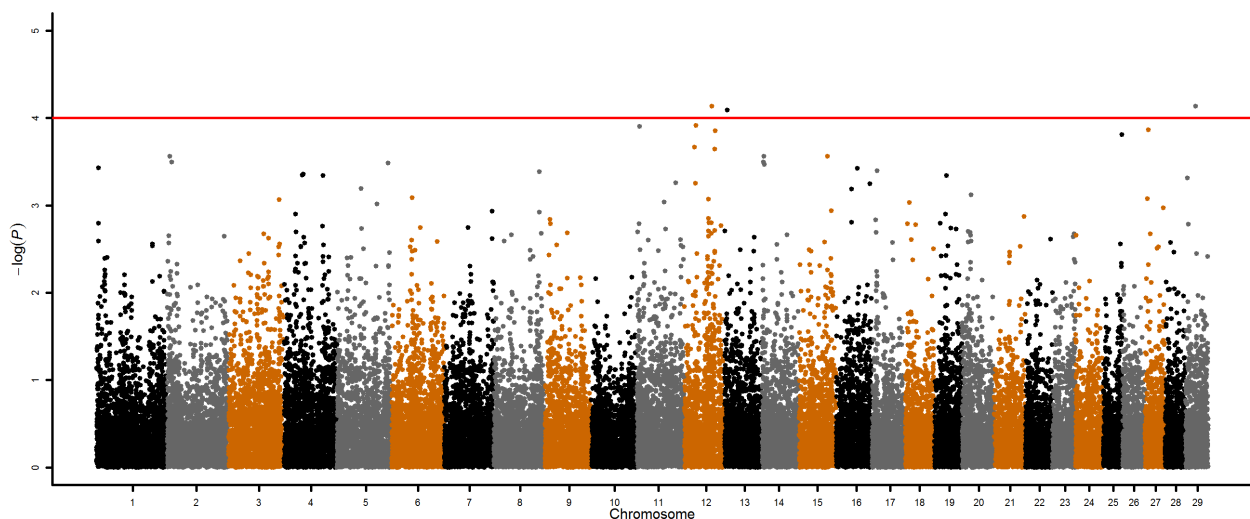


Figure 1: *Manhattan plot de los resultados ($\log(p\text{-value})$) del estudio de asociación entre SNP y desplazamiento del abomaso en vacas de raza Frisona Holandesa. La línea roja indica el umbral de significación: $-\log(0.0001) = 4$*

y esto puede causar problemas a la hora, p.ej., de probar la significancia de la asociación o de incluir efectos sistemáticos en el análisis.

Sin embargo, hay maneras de resolver estas dificultades. La significancia estadística se puede testar comparando la homocigosis media (o la frecuencia de las ROH) en casos y controles con un test-t, o aplicando el concepto clínico de “no inferioridad” (D’Agostino et al. (2003)). Se realizan dos análisis, ROH y GWAS, y se calcula el “false discovery rate” (FDR) para los dos: la hipótesis nula es que el FDR es mayor en ROH que en GWAS (ROH inferior al GWAS), la hipótesis alternativa es que los dos métodos son equivalentes (mismo FDR medio). Los efectos sistemáticos se pueden incluir utilizando un modelo estadístico para analizar las ROH (Pollott (2012)); otra alternativa es estratificar el análisis de ROH según las clases del efecto sistemático (ej. machos/hembras), o aplicarlo a los residuos de un modelo lineal. Más detalles se encuentran en Biscarini et al. (2014a,b).

Otra opción para localizar QTLs es utilizar los resultados de técnicas de remuestreo. La Figura 3 presenta la frecuencia de inclusión de los SNP en el modelo predictivo para portadores del haplotipo BH2 en función de la posición en BTA19. Las técnicas de remuestreo permiten analizar la variabilidad alrededor del parámetro estimado y se caracterizan por tener más poder estadístico y localizar QTLs de manera más robusta e independiente de los valores p y del ensayo de frecuencia a cada locus. Los métodos

propuestos pueden ser utilizados para complementar el GWAS estándar en la localización de QTLs para caracteres de interés en el genoma de animales domésticos.

References

- Aulchenko, Y. S., De Koning, D.-J., and Haley, C. (2007). Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics*, 177(1):577–585.
- Biffani, S., Dimauro, C., Macciotta, N., Rossoni, A., Stella, A., and Biscarini, F. (2015). Predicting haplotype carriers from snp genotypes in bos taurus through linear discriminant analysis. *Genetics Selection Evolution*, 47(1):4.
- Biffani, S., Morandi, N., Locatelli, V., Pravettoni, D., Boccardo, A., Stella, A., Nicolazzi, E. L., and Biscarini, F. (2014). Adding evidence for a role of the slitrk gene family in the pathogenesis of left displacement of the abomasum in holstein-friesian dairy cows. *Livestock Science*, 167:104–109.
- Biscarini, F., Biffani, S., Morandi, N., Nicolazzi, E. L., and Stella, A. (2014a). Using runs of homozygosity to detect genomic regions associated with susceptibility to infectious and metabolic diseases in dairy cows under intensive farming conditions. *Elares*, (56):1–10.

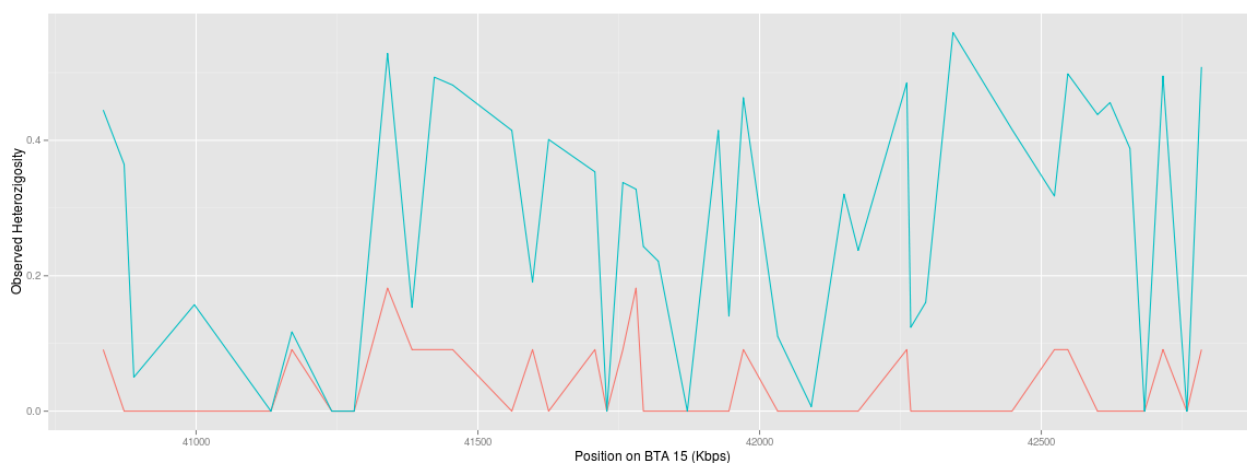


Figure 2: Valores de homocigosis a lo largo del BTA15 en casos (rojo) y controles (azul) para enfermedades reproductivas.

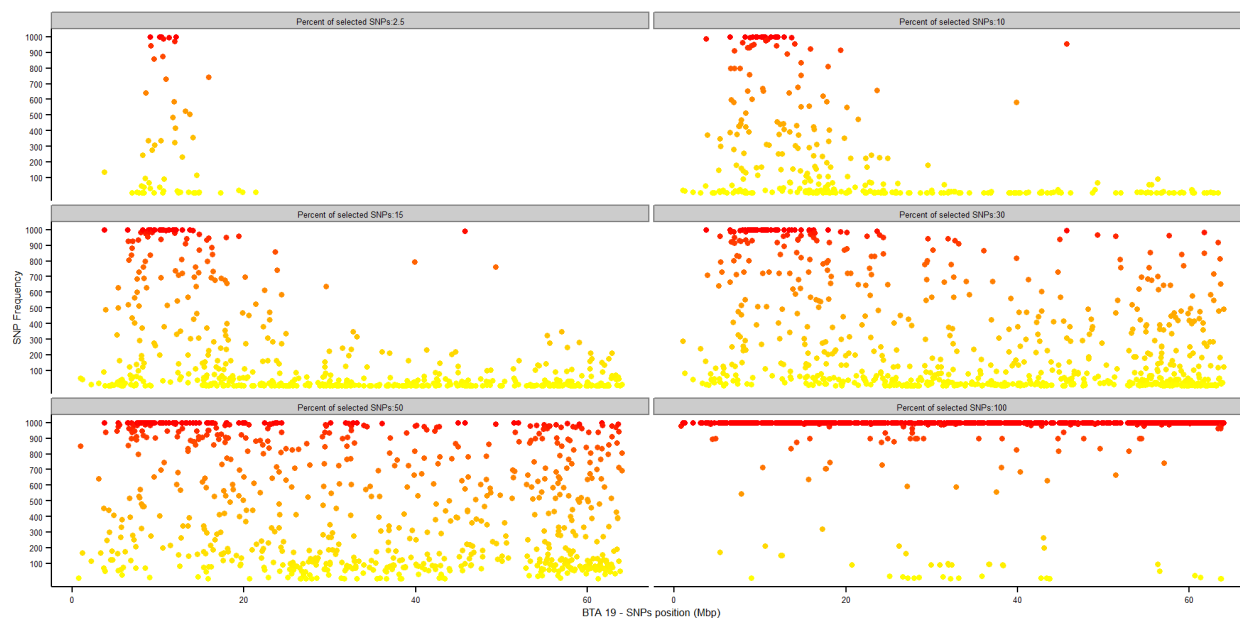


Figure 3: Frecuencia de inclusión de los SNP en el modelo predictivo (en cada de las 1000 réplicas por umbral) en función de la posición en BTA19.

- Biscarini, F., Biffani, S., Nicolazzi, E., Morandi, N., and Stella, A. (2014b). Applying runs of homozygosity to the detection of associations between genotype and phenotype in farm animals. In *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production*, number 675, pages 1–3.
- D’Agostino, R. B., Massaro, J. M., and Sullivan, L. M. (2003). Non-inferiority trials: design concepts and issues—the encounters of academic consultants in statistics. *Statistics in medicine*, 22(2):169–186.
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., and Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, 9(5):356–369.
- Pollott, G. (2012). Autozygosity by difference; a new method for locating autosomal recessive mutations. In 63rd *EAAP Book of Abstracts*, page 231.